Methodology for Evaluating a Partially Controlled Longitudinal Treatment Using Principal Stratification, With Application to a Needle Exchange Program

Constantine E. FRANGAKIS, Ronald S. BROOKMEYER, Ravi VARADHAN, Mahboobeh SAFAEIAN, David VLAHOV, and Steffanie A. STRATHDEE

We consider studies for evaluating the short-term effect of a treatment of interest on a time-to-event outcome. The studies we consider are only partially controlled in the following sense: (1) Subjects' exposure to the treatment of interest can vary over time, but this exposure is not directly controlled by the study; (2) subjects' follow-up time is not directly controlled by the study; and (3) the study directly controls another factor that can affect subjects' exposure to the treatment of interest as well as subjects' follow-up time. When factors 1 and 2 are both present in the study, evaluating the treatment of interest using standard methods, including instrumental variables, does not generally estimate treatment effects. We develop the methodology for estimating the effect of treatment 1 in this setting of partially controlled studies under explicit assumptions using the framework for principal stratification for causal inference. We illustrate our methods by a study to evaluate the efficacy of the Baltimore Needle Exchange Program to reduce the risk of human immunodeficiency virus (HIV) transmission, using data on distance of the program's sites from the subjects.

KEY WORDS: Causal inference; HIV; Needle exchange; Partially controlled studies; Potential outcomes; Principal stratification

1. INTRODUCTION

We consider studies for evaluating the short-term effect of a treatment of interest on a time-to-event outcome. For practical or ethical reasons, assume that the study design has the following features: (1) Subjects' exposure to the treatment of interest can be longitudinal, that is, can vary over time, but this exposure is not directly controlled by the study; (2) subjects' follow-up time is not directly controlled by the study; and (3) the study directly controls another factor that can affect subjects' exposure to the treatment of interest as well as subjects' follow-up time.

A motivating example is the needle exchange program (NEP) in Baltimore (ALIVE and NEP studies; Vlahov et al. 1997; Strathdee et al. 1999). In particular, a cohort of injection drug users has been enrolled and is being followed, with regular 6-month visits in which the subjects are offered clinic services, including blood tests for human immunodeficiency virus (HIV; Vlahov et al. 1991). Independently of the clinic, the NEP also operates sites in Baltimore where drug users can visit and exchange a used needle for a clean one (and receive other services such as counseling), with the hope of reducing HIV transmission. Studies for other NEP's have generally been positive, but some have shown mixed results (Bruneau, Lamothe and Franco 1997; Hagan et al. 2000). However, such studies control neither who exchanges at the NEP nor how long subjects continue the clinic visits, so comparisons of available HIV seroconversion measures between observed exchangers and nonexchangers do not generally estimate the effect of exchange. Importantly, though, the NEP staff does control placement of the NEP sites and, hence, distance of the sites from the subjects. Moreover, there have been earlier indications (e.g., Rockwell, Des Jarlais, Friedman, Rerlis, and Paone 1999) that proximity of drug users to the sites increases needle exchange. Therefore, we wish to formulate explicitly the distance as a controlled factor to provide an alternative evaluation of the NEP's effect on HIV transmission.

When there is a single partially controlled factor, the structure of studies with features 1 and 3 in general share some aspects with the more standard method of instrumental variables (IV's) in some earlier studies for other evaluations. For example, Card (1986) used families' proximity to colleges to evaluate the impact that attending college can have on later income, and McClellan, McNeil, and Newhouse (1994) used elders' proximity to hospitals to assess whether intensive treatment after myocardial infarction reduces mortality. More recently, there has also been increasing interest in using IV's to estimate the effects of treatment received in trials where some subjects do not comply with their assigned treatment (e.g., Sommer and Zeger 1991; Robins and Greenland 1994; Baker and Lindeman 1994, 2001; Imbens and Rubin 1994, 1997; Angrist and Imbens 1995; Angrist, Imbens, and Rubin 1996; Baker, Lindeman, and Kramer 2001).

Evaluation of the effect of a treatment on an outcome is more demanding in studies where that treatment as well as other factors, such as observation of the outcome, are only partially controlled. Existing frameworks for time-dependent factors such as those discussed in Robins, Greenland, and Hu (1999); Hernan, Brumback, and Robins (2000), and Murphy et al. (2001) are appropriate for their assumptions and goals. Our framework is fundamentally different from such existing frameworks in terms of the partial control for the factors we consider and in terms of the goals we have (for a comparison of frameworks, see Frangakis and Rubin 2002). In particular, in studies where such partially controlled factors as (1) and (2) given previously can interact, standard methods, including standard IV's, are not adequate to evaluate the treatment of interest (Frangakis and Rubin 1999). Moreover, in such studies, more flexible methods under more plausible assumptions have been limited to single-time treatments and all-or-none levels of the controlled factor (Frangakis and Rubin 1997, 1999; Baker 1998, 2000).

Constantine E. Frangakis is Assistant Professor and Ronald S. Brookmeyer is Professor in the Department of Biostatistics, Steffanie A. Strathdee is Associate Professor in the Department of Epidemiology, Ravi Varadhan and Mahboobeh Safaeian are doctoral students in the Departments of Biostatistics and Epidemiology, Bloomberg School of Public Health of the Johns Hopkins University, Baltimore, MD 21205. David Vlahov is Director, Center for Urban Epidemiologic Studies, New York Academy of Medicine, New York, NY. The authors thank the editor, an associate editor, three anonymous reviewers, Joshua Angrist, Phil Dawid, Donald Rubin, and Daniel Scharfstein for valuable comments. The work was supported in part by NIH (NEI) grant RO1 EY 014314-01.

Journal of the American Statistical Association
???? 2003, Vol. 0, No. 0
DOI 10.1198/00

We propose a framework that evaluates transient (short-term) effects in studies with features 1–3 and that better addresses the limitations of existing methods under certain assumptions. The next section formulates the general framework, the assumptions, the data, and the causal effects that are the estimation goal. Under the general assumptions Section 3 shows identifiability of the causal effects without parametric assumptions in large samples, and Section 4 discusses a class of parametric models and estimation that is appropriate for smaller samples. In Section 5 we demonstrate our methods in the NEP study introduced previously. In Section 6 we discuss the limitations of our method and extensions.

2. FORMULATING THE GENERAL STUDY

2.1 Principal Strata of Exposure and Principal Effects on Outcome

Consider a study that begins follow-up of a cohort of subjects, at discrete time periods $t = 1, \ldots, t_{\max}$, to investigate a discrete time-to-event outcome. At a certain period t for subject i, define $X_{i,t}$ to be the indicator such that $\{i : X_{i,t} = 1\}$ denotes the subgroup of subjects who are still in the study but have not had the outcome by the end of period $t - 1$. At the beginning of the next period t, we distinguish between the controlled and uncontrolled factors. At that period, assume each subject i in the risk set $\{i : X_{i,t} = 1\}$ can be potentially assigned different levels of a controlled factor D (e.g., distance of clinic from subject), and consider outcomes that can be potentially observed as the controlled factor would take different levels (Neyman 1923; Rubin 1974, 1977, 1978). In particular, if subject i at time t is assigned level d of the controlled factor, then let $E_{i,t}(d)$ be the potential exposure (e.g., to the needle exchange program) that subject i will have during that period, let $Y_{i,t}(d)$ be the indicator for whether the event (e.g., HIV positive) will occur during period t or not, and let $C_{i,t}(d)$ be the indicator for whether t will be the first period the subject will not provide the outcome $Y_{i,t}(d)$ (in which case we say the person will be censored at t). Also, here we focus on studies with sufficiently many time periods so that the potential values for exposure E at each time t can have one of two levels $(0, 1)$, but where the levels can vary across time. Moreover, we allow the controlled factor D to have one of possibly multiple levels $d = 1, \ldots, d_{\max}$ at each time t. Finally, we let $H_{i,t}$ denote the observed history variables of the subject up to the beginning of period t, including the past indicators for observed outcomes, censoring, and exposure, past levels of the controlled factor, as well as covariates.

We are interested in defining and estimating the transient effect of the controlled factor D on outcome Y that is attributable to exposure E. We define this effect here as the effect of the factor D on outcome Y for subjects and at times for which the controlled factor D also affects the exposure E; a technical definition is given in Section 2.2. However, the explicitly controlled factor in the study is D, not exposure E, which is possibly also affected by the factor D. Therefore, and adopting Rubin's (1978) convention that potential outcomes in the study are written as functions only of factors that are explicitly controlled by the researcher, the potential outcome Y, as a function of the controlled factors, is formally a function only of factor D, not of exposure E. For this reason, before we define the causal effects of interest, we define the vector $S_{i,t}$ of all possible exposures, $S_{i,t} = (E_{i,t}(1), \ldots, E_{i,t}(d_{\max}))$ that subject i at time t can have at different levels of D. The vector $S_{i,t}$ is called a principal stratum of exposure, and comparisons of potential outcomes at different levels of the controlled factor D within principal strata $S_{i,t}$ are called principal effects (Frangakis and Rubin 2002).

Principal strata $S_{i,t}$ and principal effects have two general properties. First, because the stratum $S_{i,t}$ is equivalent to the function that describes how exposure is affected by the controlled factor D for subject i at time t, the stratum $S_{i,t}$ itself is not affected by the controlled factor D for that subject at that time. Second, principal effects are well-defined causal effects of the factor D on outcome Y (Frangakis and Rubin 2002). Principal strata and effects are important because, as we show in Sections 2.2 and 3, they can define the causal effect of the controlled factor D on outcome Y that is attributable to exposure E. We also show that this causal effect is estimable under certain assumptions that can be different from and often more plausible than those of more standard approaches.

By limiting focus on the preceding outcomes for each subject i as functions of the factor D separately at each time t, as opposed to jointly across all times, we treat the observed actual levels of that factor, of exposure, outcome, and censoring up to period t, as components of the history variables $H_{i,t}$ specific to subject i at time t. We do this because we wish to study the transient effect of D on outcome Y that is attributable to exposure E.

An important example for evaluating transient effects is the NEP study introduced in Section 1, and for which a preliminary formulation using principal stratification is given in Frangakis, Rubin, and Zhou (2002). In this project, the factor that the NEP's staff controlled was the NEP sites' locations and, therefore, the distance $D_{i,t}$ of the closest NEP site at time t from subject i's residence. (Note: Subjects' residence was recorded at enrollment but because, subsequently, different NEP sites were placed at different semesters, the distances $D_{i,t}$ can change over time.) If the closest NEP site is placed at distance d from subject i at semester t, then let $E_{i,t}(d)$ be the indicator for whether the person will visit and exchange at the NEP during that semester, and which we label for brevity as "exchange"; let $Y_{i,t}(d)$ be the indicator for whether the subject will become HIV positive during semester t; and let $C_{i,t}(d) = 0$ if the subject will stay in the study and have the HIV test so that $Y_{i,t}(d)$ will be observed. The principal stratum $S_{i,t} = (E_{i,t}(1), \ldots, E_{i,t}(d_{\max}))$, then, is the vector of all potential exchange behaviors of subject i at time t as a function of the distance of the NEP from the subject at that time. In the NEP distance can affect exchange $(E_{i,t}(d))$, outcome $(Y_{i,t}(d))$, and censoring $(C_{i,t}(d))$. Moreover, the time from HIV infection, for example, because of exposure to an HIV-contaminated needle, until being positive on the HIV antibody test is, for all practical purposes, less than 6 months, which is the unit of time between different measurements of data. For this reason we are interested in the transient effect that distance can have on HIV status and that is attributable to exchange within a semester.

More generally, one can set up the framework to study long-term cumulative effects, although, with partially controlled factors, that would require increasing the notation and structure in the assumptions of the next section. For simplicity, we focus here on transient effects.

2.2 Data, Assumptions, and Goal

Let $D_{i,t}$ be the actual level of the factor D of subject i during period t, and let $E_{i,t} = E_{i,t}(D_{i,t})$, $Y_{i,t} = Y_{i,t}(D_{i,t})$, and $C_{i,t} = C_{i,t}(D_{i,t})$ be the actual exposure, outcome, and indicator for whether that outcome is censored, respectively, in the study. For the actual outcomes for time to event and censoring in the study, the values $Y_{i,t} = 1$ and $C_{i,t} = 1$ are absorbing states for the risk set in the sense that if either occurs during period t, then the subject will not be in the risk set $\{i' : X_{i',t'} = 1\}$ for $t' > t$. Assume, then, that for each subject i we observe the indicator $X_{i,t}$ for whether the subject is in the risk set, that is, has not had the outcome and has not been censored by the beginning of period t, and, if the subject is in that risk set, then the observed data during period t also include the following: the levels of the factors $D_{i,t}$ and $E_{i,t}$; the censoring indicator $C_{i,t}$ for the outcome $Y_{i,t}$; the outcome $Y_{i,t}$ if the subject is not censored during period t; and history variables $H_{i,t}$, which can include the levels of the factors $D_{i,t'}$ and $E_{i,t'}$, as well as covariates, for times $t' < t$. (Note: The assumption that $D_{i,t}$ and $E_{i,t}$ are observed even at the time of possible censoring of the person's $Y_{i,t}$ can be relaxed; it is made here because it is true in the NEP example, owing to the different sources supplying information in that study; see Vlahov et al. 1997.)

We consider the subjects i as a random sample from a large population to which we wish to generalize, and assume that the following conditions hold.

1. Sequential Ignorability of Assignment to the Controlled Factor.

$$\left(S_{i,t}, \left\{Y_{i,t}(d), C_{i,t}(d),\ d = 1, \ldots, d_{\max}\right\}\right) \perp\!\!\!\perp D_{i,t} \mid H_{i,t}, X_{i,t} = 1 \quad \text{for times } t \geq 1,$$

where Dawid's (1979) symbol $\perp\!\!\!\perp$ denotes independence. The assumption expresses that, among subjects still in the risk set by time t, the study assigns levels $D_{i,t}$ of the controlled factor at random with probabilities that can depend on the observed history variables $H_{i,t}$. Sequential ignorability for a controlled factor has been considered earlier, for example, by Rubin (1978), Robins (1986), Robins et al. (1999), Scharfstein, Rotnitzky, and Robins (1999), and Murphy et al. (2001) but in settings with fundamentally different structure from ours in the degree of control for exposure and censoring behavior.

In the NEP example, where the controlled factor was the distance $D_{i,t}$ of the closest NEP site from each drug user, within areas of past high risk for HIV transmission, the NEP sites were placed in a nonsystematic way. This makes sequential ignorability more plausible when we include in $H_{i,t}$ main variables, such as sexual and drug use behavior, that can be confounders with $D_{i,t}$ in the sense that they cluster in some of Baltimore's areas. More generally, sequential ignorability is an explicit way to express conditional comparability of subjects at different levels of the controlled factor D.

2. Multilevel Monotonicity of Exposure. We assume that increasing levels of the factor D provide decreasing encouragement for exposure, or, formally, that $E_{i,t}(d') \leq E_{i,t}(d)$ if $d' > d$. Multilevel monotonicity is a generalization of the monotonicity assumption of Imbens and Rubin (1994) for a binary controlled factor. For a single partially controlled factor, in particular, with no outcome censoring, the previous assumption was considered by Angrist and Imbens (1995) for estimation of linear models in instrumental variables and by Baker et al. (2001).

Under multilevel monotonicity in our framework, subject i's principal stratum $S_{i,t}$ is one of $d_{\max} + 1$ levels, say, $0, 1, \ldots, d_{\max}$, and is equivalent to the subject-specific threshold (maximum) level that the controlled factor can be, above which subject i at time t would not become exposed to treatment E. This ordered structure is represented in Figure 1(a).

In the NEP example monotonicity formally reflects that if a subject at time t is one who would not visit the NEP to exchange needles if the NEP were at a distance d, then the subject would also not visit the NEP to exchange if the NEP were at a longer distance, which is a generally plausible condition (e.g., Rockwell et al. 1999). Moreover, under monotonicity, a statement about the principal stratum $S_{i,t} = (E_{i,t}(1), \ldots, E_{i,t}(d_{\max}))$ of subject i, that is, about all $d_{\max}$ exposure behaviors of the subject as a function of distance, is equivalent to a statement about the closest distance we can place the NEP beyond which that subject will not exchange needles at it during period t. For this reason we will henceforth be using the notation $S_{i,t} = d$, $d = 0, \ldots, d_{\max}$, for that closest distance to indicate the principal stratum of that subject at that time.

Note that, even under monotonicity, a subject's principal stratum $S_{i,t}$ generally is not fully observed and is different from the observed strata defined by $(D_{i,t}, E_{i,t})$. For example, as Figure 1, (a) and (c), shows, subjects who exchange when being at the closest distance from the NEP ($D_{i,t} = 1, E_{i,t} = 1$) are a mixture of all subjects with principal stratum $S_{i,t} \geq 1$.

3. Compound Exclusion Restriction. A subject with $E_{i,t}(d) = E_{i,t}(d')$ at time t is, by definition, someone whose exposure behavior E would be the same if the controlled factor D were set at level d or d'. For this reason we assume that assigning such a subject to either level d or d' of D would not change the subject's outcome, Y, or the behavior, C, of censoring of the outcome. That is,

$$\text{if } E_{i,t}(d) = E_{i,t}(d') \quad \text{then } Y_{i,t}(d) = Y_{i,t}(d') \text{ and } C_{i,t}(d) = C_{i,t}(d').$$

For a particular time Figure 1(b) displays the pattern that compound exclusion implies for the average, say, $\bar{y}(S, D)$, of the potential outcomes when principal stratum S is at level D of the controlled factor (the time index is omitted); average censoring rates follow an analogous pattern.

Special cases of the preceding compound exclusion restriction have been explored for no censoring, for example, by Imbens and Rubin (1994) and Baker et al. (2001), and with censoring but binary controlled factor, for example, by Frangakis and Rubin (1997, 1999), Baker (1998), and Barnard, Frangakis, Hill, and Rubin (2002).

More generally, the principal strata and compound exclusion restriction help define causal effects of interest. In particular, at time t, for subjects in the two principal strata, $S_{i,t} = 0$ and $S_{i,t} = d_{\max}$, whose exposure is the same for all levels of the controlled factor D [Fig. 1(a)], there is no causal effect of D on Y [Fig. 1(b)]. For subjects in any one of the remaining principal strata, $S_{i,t} = d$, $1 \leq d \leq d_{\max} - 1$, consider the comparison between two proportions: the proportion of subjects who get the

(a) (b)

Principal stratum means	what are all exposure behaviors as functions of the controlled factor D, E(S, D)					fractions of people with Y=1, given S, as functions of the controlled factor D, $\bar{y}(S, D)$				
S	$D = d_{max}$	$D = d_{max} - 1$	$\ldots$	$D = 2$	$D = 1$	$D = d_{max}$	$D = d_{max} - 1$	$\cdots$	$D = 2$	$D = 1$
$S = d_{max}$	1	1	$\cdots$	1	1	$\bar{y}(d_{max}, d_{max})$	·	$\cdots$	·	
$S = d_{max} - 1$	0	1	$\cdots$	1	1	$\bar{y}(d_{max} - 1, d_{max})$	$\bar{y}(d_{max} - 1, 1)$	$\cdots$	·	$\bar{y}(d_{max} - 1, 1)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$					
$S = 2$	0	0	$\cdots$	1	1	$\bar{y}(2, d_{max})$	$\bar{y}(2, d_{max})$	$\cdots$	$\bar{y}(2, 1)$	$\bar{y}(2, 1)$
$S = 1$	0	0	$\cdots$	0	1	$\bar{y}(1, d_{max})$	·	$\cdots$	$\bar{y}(1, d_{max})$	$\bar{y}(1, 1)$
$S = 0$	0	0	$\cdots$	0	0	$\bar{y}(0, d_{max})$	·	$\cdots$	·	·

(c)

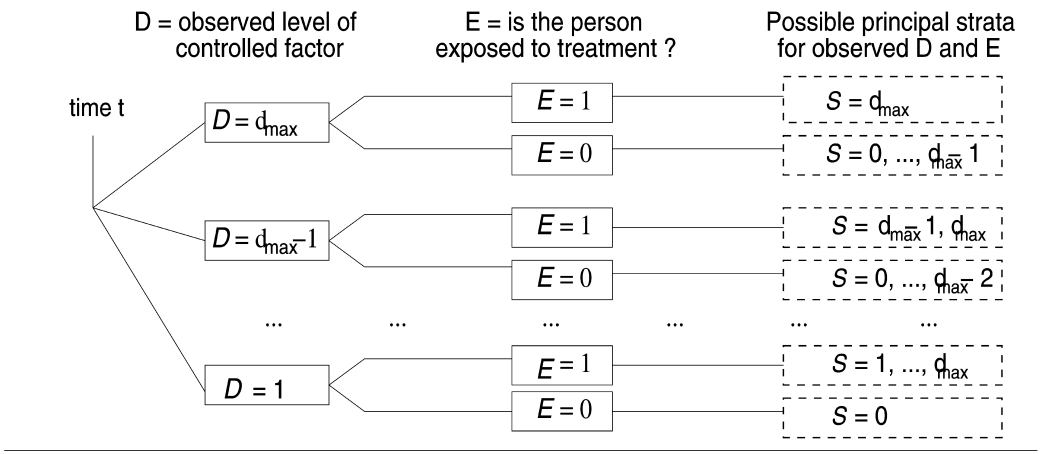

Figure 1. Principal Stratification at a Particular Time (time index is omitted). Relations in (a) follow by multilevel monotonicity, and in (b) by compound exclusion restriction.

event if assigned $D = 1$ versus the proportion of subjects who get the event if assigned $D = d_{\max}$, that is,

$$\begin{aligned} &\text{pr}(Y_{i,t}(1) = 1 \mid H_{i,t}, X_{i,t} = 1, S_{i,t} = d) \\ &\text{versus} \\ &\text{pr}(Y_{i,t}(d_{\max}) = 1 \mid H_{i,t}, X_{i,t} = 1, S_{i,t} = d). \end{aligned} \tag{2.1}$$

Such comparisons of potential outcomes within principal strata, called principal effects, date back at least to Imbens and Rubin (1994) for randomized studies with noncompliance, and were formulated by Frangakis and Rubin (2002) to principal strata for general partially controlled variables. Before the general definition of such estimands, standard definitions of treatment effects in settings with both a controlled and a partially controlled factor were comparisons between populations of different groups of subjects ("net-treatment effects," e.g., Cochran 1957; Rosenbaum 1983), and so were not generally causal effects (see, for example, Rosenbaum 1983; Frangakis and Rubin 2002). The importance of principal effects in (2.1) is that both proportions in (2.1) are conditional on the same group of subjects, so that their comparison is a well-defined causal effect (Frangakis and Rubin 2002). Moreover, every subject in the top expression of (2.1) gets exposure to treatment E, whereas no subject in the bottom expression of (2.1) gets exposure to treatment E [Fig. 1, (a) and (b), for $S_{i,t} = 1, \ldots, d_{\max} - 1$]. Therefore, a principal effect of the form (2.1) better quantifies an effect of the controlled factor D on outcome Y that is attributable to exposure E. For this reason we set (2.1) to be our main goal for estimation.

In the NEP compound exclusion means that two different distances of the NEP location from a specific subject might result in different HIV status Y or censoring of that status (e.g., by discontinuation of follow-up), but only if those two distances would result in different needle exchange (and visiting) behavior at the NEP, an assumption that is expected to be approximately true. In this case a principal effect of the form (2.1) quantifies an effect that distance from the NEP can have on HIV transmission and that is attributable to exchange of needles at the NEP. This effect is of main interest in the application.

4. Latent Ignorability of the Censoring Mechanism. Because the principal stratum $S_{i,t}$ is a characteristic of the subject, effectively representing the willingness to become exposed to treatment, $S_{i,t}$ can be associated with both the potential outcome and the censoring of the outcome. In the NEP example that would mean that subjects willing to exchange at different distances can have different risk for HIV and/or different willingness to stay in the study. More generally, then, if we knew the principal stratum $S_{i,t}$ for all subjects, we should first stratify subjects by it, before connecting outcomes of censored subjects (i.e., without measured outcomes) to uncensored subjects (i.e., with measured outcomes). An assumption that formalizes this is

$$Y_{i,t}(d) \perp\!\!\!\perp C_{i,t}(d) \mid D_{i,t}, H_{i,t}, X_{i,t} = 1, \text{ and } S_{i,t}$$

for times $t \geq 1$ and each level d of the controlled factor. (Note: By definition, observed exposure $E_{i,t}$ is a function of $D_{i,t}$ and $S_{i,t}$, so the right side of the previous expression also includes $E_{i,t}$.) The preceding assumption extends the latent ignorability for single time points introduced by Frangakis and Rubin (1999) for randomized trials with noncompliance.

Because neither exposure nor censoring is controlled, of course, any set of assumptions, including the previous four, is not guaranteed to be absolutely correct. Therefore, the set of assumptions is best judged not by whether it is certain to be absolutely correct, but in comparison to other existing approaches, on the inferences and the sensitivity analyses it can produce for estimating the effects (2.1). Specifically, a framework that allows the more general condition, "A" = "the mechanism of censoring C can depend on the latent principal stratum S," is richer and so can produce (a) inferences for (2.1) that are better than inferences that do not allow "A" and (b) sensitivity analyses around inferences that allow "A" and, thus, are more informative than sensitivity analyses around inferences that do not allow "A" (for a related discussion, see Frangakis and Rubin 1999, sec. 4.2). To our knowledge, the framework using latent ignorability is the first in this longitudinal setting that both allows "A" and also allows identifiability of the causal effects (2.1), as discussed in the next section. In Section 4 we discuss parametric models for our set of assumptions, although study of sensitivity to models around those assumptions and alternative assumptions that would allow "A" is also of interest.

3. IDENTIFIABILITY OF PRINCIPAL CAUSAL EFFECTS

We show that, under the assumptions of the previous section, the probabilities in (2.1), and, hence, comparisons among them, are identifiable with no parametric assumptions. Without loss of generality, for this section, suppose we are already within an observed stratum of the subjects at risk at the beginning of a specific period t and already within an observed stratum defined by history $H_{i,t}$. The first three of the following six expressions define the notation for the proportions of principal strata, of no censoring, and of positive outcome within principal strata and when the subjects get assigned $D = d$. The last three expressions define the notation for the directly estimable proportions of observed exposure within strata $D = d$, of no censoring within strata $D = d$ and exposure $E = e$, and of positive outcome within strata $(D = d, E = e)$ and among uncensored subjects $(C = 0)$,

$$\begin{aligned} \bar{s}(d) &:= \text{pr}(S_i = d), \\ \bar{c}(d', d) &:= \text{pr}(C_i(d) = 0 \mid S_i = d'), \\ \bar{y}(d', d) &:= \text{pr}(Y_i(d) = 1 \mid S_i = d'), \\ \bar{e}^{\text{obs}}(d) &:= \text{pr}(E_i = 1 \mid D_i = d), \\ \bar{c}^{\text{obs}}(e, d) &:= \text{pr}(C_i = 0 \mid D_i = d, E_i = e), \\ \bar{y}^{\text{obs}}(e, d) &:= \text{pr}(Y_i = 1 \mid D_i = d, E_i = e, C_i = 0), \end{aligned}$$

where the indexing for time t is omitted. Assuming the probabilities in the first two lines are in $(0, 1)$, we derive the probabilities in the left column as a function of those in the right column.

Proportions of Principal Strata. Note that the observed stratum $(D = d_{\max}, E = 1)$ contains only the principal stratum $S = d_{\max}$. By multilevel monotonicity an observed stratum $(D = d, E = 1)$ of exposed subjects at level $D = d$ is the mixture [Fig. 1, (a) and (c)] of the principal strata, $S = d, \ldots, d_{\max}$, so

$$\bar{e}^{\text{obs}}(d_{\max}) = \bar{s}(d_{\max}), \qquad \bar{e}^{\text{obs}}(d) = \bar{s}(d_{\max}) + \cdots + \bar{s}(d),$$

and so

$$\bar{s}(d) = \bar{e}^{\text{obs}}(d) - \bar{e}^{\text{obs}}(d+1).$$

Proportions of Uncensored Subjects Within Principal Strata. Because the observed stratum $(D = d_{\max}, E = 1)$ contains only subjects with $S = d_{\max}$, we have that $\bar{c}^{\text{obs}}(1, d_{\max}) = \bar{c}(d_{\max}, d_{\max})$, which equals $\bar{c}(d_{\max}, 1)$ by compound exclusion. The noncensoring proportions in the remaining observed strata $(D = d, E = 1)$, for $1 \le d \le d_{\max} - 1$, are mixtures across principal strata. Let $\bar{s}_{(+)}(d) = \bar{s}(d) + \cdots + \bar{s}(d_{\max})$. Then

$$\begin{aligned}\bar{c}^{\text{obs}}(1, d) &= \frac{\bar{s}(d_{\max})\bar{c}(d_{\max}, d) + \cdots + \bar{s}(d)\bar{c}(d, d)}{\bar{s}_{(+)}(d)} \\ &= \frac{\bar{s}(d_{\max})\bar{c}(d_{\max}, 1) + \cdots + \bar{s}(d)\bar{c}(d, 1)}{\bar{s}_{(+)}(d)},\end{aligned}$$

where the last equality follows by compound exclusion. Considering the last equality for $d + 1$ and using induction in the numerators, we get

$$\bar{c}(d, 1) = \frac{\bar{s}_{(+)}(d)\bar{c}^{\text{obs}}(1, d) - \bar{s}_{(+)}(d+1)\bar{c}^{\text{obs}}(1, d+1)}{\bar{s}(d)}.$$

Working similarly with induction for strata $(D = d, E = 0)$, we obtain that $\bar{c}(0, d_{\max}) = \bar{c}^{\text{obs}}(0, 1)$, and, for the remaining principal strata $S = d$, for $1 \le d \le d_{\max} - 1$, that

$$\begin{aligned}&\bar{c}(d, d_{\max}) \\ &\quad = \frac{\{1 - \bar{s}_{(+)}(d+1)\}\bar{c}^{\text{obs}}(0, d+1) - \{1 - \bar{s}_{(+)}(d)\}\bar{c}^{\text{obs}}(0, d)}{\bar{s}(d)}.\end{aligned}$$

Proportions of Potential Outcomes Within Principal Strata. As with censoring rates, because the observed stratum $(D = d_{\max}, E = 1)$ contains only the principal stratum $S = d_{\max}$, we have $\bar{y}^{\text{obs}}(1, d_{\max}) = \bar{y}(d_{\max}, d_{\max}) = \bar{y}(d_{\max}, 1)$, where the first equality follows by latent ignorability, and the second by compound exclusion. For the remaining observed strata $(D = d, E = 1)$, $1 \le d \le d_{\max} - 1$, the proportion for the outcome that is observable, $\bar{y}^{\text{obs}}(1, d)$, is a mixture over principal strata S:

$$\begin{aligned}\bar{y}^{\text{obs}}(1, d) = E\big\{&E(Y = 1 \mid D = d, E = 1, S) \\ &\mid D = d, E = 1, C = 0\big\},\end{aligned}$$

where, in the inner expectation, $C = 0$ is canceled by latent ignorability. After some algebra the previous expression can be written as

$$\begin{aligned}&\bar{y}^{\text{obs}}(1, d) \\ &\quad = \frac{\bar{s}(d)\bar{c}(d, 1)\bar{y}(d, 1) + \cdots + \bar{s}(d_{\max})\bar{c}(d_{\max}, 1)\bar{y}(d_{\max}, 1)}{\bar{c}_{+}(d)},\end{aligned}$$

where $\bar{c}_{+}(d) = \bar{s}(d)\bar{c}(d, 1) + \cdots + \bar{s}(d_{\max})\bar{c}(d_{\max}, 1)$. By considering the preceding expression also for $d + 1$ and using induction, we obtain the outcome proportion within principal strata as a function of quantities that we have already shown are estimable:

$$\bar{y}(d, 1) = \frac{\bar{c}_{+}(d)\bar{y}^{\text{obs}}(1, d) - \bar{c}_{+}(d+1)\bar{y}^{\text{obs}}(1, d+1)}{\bar{s}(d)\bar{c}(d, 1)}. \quad (3.1)$$

With an analogous argument, with induction for the outcome probability on the strata $(D = d,\ E = 0)$, we obtain that $\bar{y}(0, d_{\max}) = \bar{y}^{\text{obs}}(0, 1)$, and, for the remaining principal strata $S = d$, for $1 \le d \le d_{\max} - 1$, that

$$\bar{y}(d, d_{\max}) = \frac{\bar{c}_{(-)}(d+1)\bar{y}^{\text{obs}}(0, d+1) - \bar{c}_{(-)}(d)\bar{y}^{\text{obs}}(0, d)}{\bar{s}(d)\bar{c}(d, d_{\max})}, \quad (3.2)$$

where $\bar{c}_{(-)}(d) = \bar{s}(0)\bar{c}(0, d_{\max}) + \cdots + \bar{s}(d-1)\bar{c}(d-1, d_{\max})$. Expressions (3.1) and (3.2) establish identifiability, using observed data, of the outcome probabilities within principal strata and, therefore, identifiability of the principal causal effects (2.1).

There are two more general notes regarding the principal effects. First, note that their expressions (3.1) and (3.2) involve the censoring mechanism $\bar{c}(d', d)$ (estimable as shown earlier), which, therefore, is not ignorable in the sense of Rubin (1976). Second, we can write (3.1) and (3.2) as

$$\begin{aligned}\bar{y}(d, 1) &= \frac{\frac{\delta}{\delta(d)}\bar{c}_{(+)}(d)\bar{y}^{\text{obs}}(1, d)}{\frac{\delta}{\delta(d)}\bar{c}_{(+)}(d)}, \\ \bar{y}(d, d_{\max}) &= \frac{\frac{\delta}{\delta(d)}\bar{c}_{(-)}(d)\bar{y}^{\text{obs}}(0, d)}{\frac{\delta}{\delta(d)}\bar{c}_{(-)}(d)},\end{aligned} \quad (3.3)$$

where $\delta/\delta(d)$ denotes the finite-difference derivative with respect to d. The last two expressions, although ratios of differences, are not those used in standard IV models with additive error terms (e.g., Wald 1940; Bowden and Turkington 1984). Therefore, the causal effects (2.1) are identifiable as before, but not using IV's.

4. ESTIMATION USING PARAMETRIC MODELS

Expressions (3.3) are useful for establishing identifiability of the causal effects without parametric assumptions in large samples at any time period and given past history, but are not directly useful for practical estimation of the effects with many time periods and with continuous covariates. In such cases parametric models can smooth the relation between the variables across the many strata.

A useful class of parametric models is obtained when we place models on the distributions in the left column of the expression at the beginning of Section 3, which are of direct interest. Allowing now, explicitly, dependence on time and past covariates, we can model the distribution of the ordinal principal strata by the proportional odds model, where

$$\begin{aligned}&\operatorname{logit}\operatorname{pr}(S_{i,t} \ge d \mid H_{i,t} = h, X_{i,t} = 1, \beta^{(S)}) \\ &\qquad = \beta^{(S)}_{(d)} + \operatorname{link}^{(S)}(h, t)\beta^{(S)}_{(h)}, \quad (4.1)\end{aligned}$$

where $\operatorname{link}^{(S)}(h, t)$ is a link function, $\beta^{(S)}_{(d)} \ge \beta^{(S)}_{(d')}$ for $1 \le d < d' \le d_{\max}$, $\operatorname{logit}(\cdot) = \log(\cdot/(1 - \cdot))$, and where the probability for $S_{i,t} = 0$ is determined by the previous ones.

We can model the target probability of the event $Y = 1$ in (2.1) and the analogous probability of censoring $C = 1$ of that event as functions of the factor D and given principal strata, respectively, by the logistic models:

$$\begin{aligned}&\operatorname{logit}\operatorname{pr}(Y_{i,t}(d) = 1 \mid H_{i,t} = h, X_{i,t} = 1, S_{i,t} = d', \beta^{(Y)}) \\ &\qquad = \operatorname{link}^{(Y)}(d, d', h, t)\beta^{(Y)}, \quad (4.2)\end{aligned}$$

$$\operatorname{logit}\operatorname{pr}(C_{i,t}(d)=1 \mid H_{i,t}=h, X_{i,t}=1, S_{i,t}=d', \beta^{(C)}) = \operatorname{link}^{(C)}(d,d',h,t)\beta^{(C)}, \quad (4.3)$$

where we require that $\operatorname{link}^{(Y)}(d,d',h,t)$ and $\operatorname{link}^{(C)}(d,d',h,t)$ be functions that satisfy the implications of the compound exclusion restriction in the population. An example is the function $\operatorname{link}^{(Y)}(d,d',h,t) = \operatorname{link}^{(C)}(d,d',h,t) = [1,h,t,d', E_{(S,D)}(d',d)]$, where $E_{(S,D)}(d',d)$ is the matrix of Figure 1(a), with (d',d) entry equal to the exposure that principal stratum d' has when assigned level d of the factor D. Moreover, the parameters $\beta^{(S)}$, $\beta^{(Y)}$, and $\beta^{(Y)}$ can be tied together by additional plausible conditions, if warranted, to increase precision of estimation.

Using models (4.1)–(4.3), and noting that the actual exposure data $E_{i,t}$ are functions of the person's principal stratum $S_{i,t}$ and the factor level $D_{i,t}$, we can obtain the likelihood of observing the data of person i at period t as if these data also included the current principal stratum $S_{i,t}=d'$, and conditionally on the risk set $X_{i,t}=1$, history $H_{i,t}=h$, and level $D_{i,t}=d$, as

$$\begin{aligned} & l(d',d,y,c,h,t;\beta) \\ & := \operatorname{pr}\big(S_{i,t}=d' \mid H_{i,t}=h, X_{i,t}=1, \beta^{(S)}\big) \\ & \quad \times \big\{\operatorname{pr}(Y_{i,t}(d)=y \mid H_{i,t}=h, X_{i,t}=1, S_{i,t}=d', \beta^{(Y)})\big\}^{(1-c)} \\ & \quad \times \operatorname{pr}\big(C_{i,t}(d)=c \mid H_{i,t}=h, X_{i,t}=1, S_{i,t}=d', \beta^{(C)}\big), \end{aligned}$$

where the last product follows by latent ignorability and where $\beta := (\beta^{(S)}, \beta^{(Y)}, \beta^{(C)})$. Now let $\mathcal{S}(D_{i,t}, E_{i,t})$ be the set of possible principal strata that subject i can belong to at time t, as a function of the observed factor level $D_{i,t}$ and observed exposure $E_{i,t}$; that is, $\mathcal{S}(D_{i,t}, E_{i,t}) = \{d : d \geq D_{i,t}\}$ if $E_{i,t}=1$, and $\mathcal{S}(D_{i,t}, E_{i,t}) = \{d : d < D_{i,t}\}$ if $E_{i,t}=0$ [Fig. 1(c)]. Then, using the previous likelihood function, we can obtain a partial likelihood function $L(\beta)$ as the product of the likelihood of observed data over all subjects in the risk set at period t, and over all periods t, as

$$L(\beta) = \prod_t \prod_{i:X_{i,t}=1} \sum_{s \in \mathcal{S}(D_{i,t},E_{i,t})} l(s, D_{i,t}, Y_{i,t}, R_{i,t}, X_{i,t}, t; \beta).$$

The function $L(\beta)$ is a partial likelihood, not in the usual sense of Cox's (Cox and Oakes 1984) partial likelihood, but in the sense that $L(\beta)$ omits the distributions of the variables in $H_{i,t}$ that are not past observed outcome, censoring, or exposure indicators. Nevertheless, as with partial likelihood in other settings (e.g., Cox and Oakes 1984; Robins et al. 1999), the estimator that maximizes $L(\beta)$ enjoys properties analogous to those of the usual maximum likelihood estimators (MLE's), in the sense that, in large samples, its distribution is approximately normal, centered around the true parameter vector, and with variance estimated consistently by the inverse of the negative second derivative of $\log L(\beta)$, evaluated at the MLE. Then a comparison between the probabilities in (2.1) can be estimated by expressing them in terms of the models (4.2) and estimating them by the MLE.

We have developed a program for maximizing $L(\beta)$ for general levels of principal strata S and periods t, using an EM algorithm (Dempster, Laird, and Rubin 1977) that treats the partially observed principal strata as incomplete data. The Hessian matrix is obtained numerically at the MLE. The program was checked using limited preliminary simulations and is available from the authors.

5. EXAMPLE ON NEEDLE EXCHANGE

We return to the NEP study, introduced in Section 1, and the goal is to evaluate the NEP's impact on HIV seroconversion. In Section 5.2 we present an evaluation based on a standard method, and in Section 5.3, an evaluation based on the new method. First, we discuss some additional background.

5.1 Background

The cohort in this evaluation consists of the 1,170 subjects who were HIV negative in 1994 and for whom residence information was available at that time. Since then, the average follow-up was 9 semesters, during which time 54 subjects tested HIV positive (5 per 1,000 person-semesters), and the overall rate of subjects' needle exchange visits was 14 per 100 person-semesters. In addition, before starting to place NEP sites in 1994, the NEP staff requested a set of 25 baseline covariates for each subject in the study, measuring various aspects of sexual and drug use behaviors. Directly fitting all these covariates in a model (4.1)–(4.3) is not possible, because many covariate values were missing across subjects, and also because of the proportionally small number of HIV cases.

To address the missing values for the covariates only, but to limit further complexity, we used multiple imputation (Rubin 1987, 1996). Specifically, we fitted a standard Bayesian general location model on the contingency table of the discrete covariates and on the continuous covariates conditionally on that contingency table, assuming missing covariate values are missing at random and fitting all the second-order interactions, using software by Schafer (1998). We ran five independent dataaugmentation chains, and, after they mixed, as judged by the potential scale reduction criterion (Gelman and Rubin 1992), the covariate missing values were imputed from their joint posterior predictive distribution conditionally on the observed covariate values, giving five complete-data covariate matrices. Each such matrix was then matched to the data matrix for the longitudinal distance D, exchange E, HIV status Y, and censoring C of Y, creating five sets of data, each of whose structure is as described in the first paragraph of Section 2.2. Each of the five datasets was analyzed as described in Sections 4 and 5, and the analyses were combined at the end by the general multiple imputation rules (Rubin 1987).

To handle the small number of HIV cases up to the calendar time of this analysis, we had to limit distance levels to two, $d_{\max}$ for more than 3 miles (far) and $d=1$ for 3 miles or less (close) of the NEP from each subject. To address the multitude of covariates, after the imputations described previously, we transformed them into their principal components. The first component, as expected, gave positive weights to those levels of the covariates that were positive on risky sexual and drug use behaviors. For this reason, we used the first principal component of subject i, denoted by B_i, as a summary index of an observed baseline risk behavior for HIV transmission in our models. The uncertainty in the final results was almost entirely from within imputations, which suggests that results are robust to the preceding preparatory computations.

Table 1. Percentage (%) of Subjects Exchanging at the NEP as a Function of Distance From the NEP, at Each Semester, and Stratified by Above (high) and Below (low) the Median Baseline Risk Score

	Semester (t)												
Distance (D)	1	2	3	4	5	6	7	8	9	10	11	12	Average
(a) For subjects with low baseline risk score:													
Far	4	5	6	6	6	5	6	4	6	5	7	2	5
Close	8	7	7	8	7	12	12	11	13	13	12	9	10
(b) For subjects with high baseline risk score:													
Far	16	18	21	19	18	16	14	14	12	14	7	5	14
Close	24	24	24	19	18	19	19	20	21	21	20	16	20

5.2 Evaluation With Observed Stratification

Table 1 gives direct estimates of the percentage of at-risk of subjects who exchange at the NEP, that is, who visit an NEP at least once in the semester, stratified by whether the NEP is far or close to them, and stratified by whether their baseline risk score B is above or below its median, m_B. These descriptive measures show that subjects with high observed baseline risk exchange considerably more than the others. Moreover, subjects closer to the NEP also exchange more frequently than subjects farther from the NEP, a result that confirms an earlier finding by Rockwell et al. (1999).

To quantify these relations while also modeling past exchange, for subjects concurrently at the risk set, we fit the model

$$\begin{aligned}&\operatorname{logit}\operatorname{pr}(E_{i,t}=1 \mid D_{i,t}, H_{i,t}, X_{i,t}=1)\\ &\quad=\alpha_t^{(E)}+[D_{i,t}\cdot 1(B_i<m_B), D_{i,t}\cdot 1(B_i\geq m_B),\\ &\qquad B_i, E_{i,t-1}, D_{i,t-1}]'\alpha^{(E)}\end{aligned}$$

by conditional logistic regression using its partial likelihood. In this and later models, the baseline risk score and first past lags of exchange and distance are included in the model as history variables to make the strata more comparable in past behaviors. With this model we obtain that, for subjects with low baseline risk, there is a 2.2 odds ratio for exchanging at close compared to far distance from the NEP (95% CI: 1.64, 2.96), and for subjects with high baseline risk, these odds are 1.70 (95% CI: 1.29, 2.41). Moreover, within a given distance $D_{i,t}$, there is a 9% increase (95% CI: 5%, 13%) in the odds of exchanging needles associated with an increase of 1 standard deviation in the baseline risk B. In addition, the odds of exchanging were 26-fold higher (95% CI: 22, 29) for subjects who exchanged during the previous period compared to those who did not, given distance and baseline risk score.

Table 2. Results From the Standard Stratification Model of Section 5.2 on Exchanging Needles at the NEP and HIV Seroconversion

	Estimand	Estimate	95% CI
(a)	Odds ratio of HIV seroconversion for 1 standard deviation increase in baseline risk score B: $\exp(\alpha_{(b)}^{(Y)})$	1.17	(1.05, 1.30)
(b)	Odds ratio of HIV seroconversion for comparing exchangers versus nonexchangers, given fixed baseline risk score B: $\exp(\alpha_{(e)}^{(Y)})$	.68	(.26, 1.77)

For a standard approach to HIV seroconversion, we fitted the discrete survival model

$$\begin{aligned}&\operatorname{logit}\operatorname{pr}(Y_{i,t}=1 \mid E_{i,t}, D_{i,t}, H_{i,t}, X_{i,t}=1)\\ &\quad=\alpha_t^{(Y)}+\alpha_{(e)}^{(Y)}E_{i,t}+\alpha_{(d)}^{(Y)}D_{i,t}\\ &\qquad+\alpha_{(b)}^{(Y)}B_i+\alpha_{(e-)}^{(Y)}E_{i,t-1}+\alpha_{(d-)}^{(Y)}D_{i,t-1}.\end{aligned}$$

The two main results of interest from this approach are shown in Table 2. In particular, as would be expected, a higher baseline risk score B predicts more HIV seroconversion [Table 2(a)]. Moreover, conditionally on the other variables in that model, observed exchangers have 32% lower odds of HIV seroconversion than observed nonexchangers, although this result is not statistically significant (OR = .68; 95% CI: .26, 1.77).

The relatively wide uncertainty about the preceding difference is likely a result of the low overall seroconversion in the cohort. Moreover, for $\exp(\alpha_{(e)}^{(Y)})$ in this approach to have interpretation of causal effect as the odds ratio of HIV seroconversion attributable to exchange, three assumptions must hold: (1) Exchange can only affect HIV seroconversion in the same period; (2) given $D_{i,t}$, $H_{i,t}$, and $X_{i,t}=1$, observed exchangers and nonexchangers are comparable in their potential outcomes of HIV seroconversion; and (3) given $D_{i,t}$, $H_{i,t}$, and $X_{i,t}=1$, subjects at a time t who provide the outcome Y are comparable with subjects who do not provide the outcome. The first assumption is physiologically appropriate, and we assume it also in the framework of principal stratification. However, the second and third assumptions are not necessarily plausible because, given $D_{i,t}$, $H_{i,t}$, and $X_{i,t}=1$, exchangers and nonexchangers are different mixtures of principal strata [Fig. 1(c)]. For these reasons we also evaluated the NEP using principal stratification.

5.3 Evaluation With Principal Stratification

We now set our goal to estimate the causal effect of distance on HIV seroconversion that is the odds ratio comparison (2.1), that is, that is attributable to exchanging versus not exchanging needles, using principal stratification. To do so, we fitted the model described in Section 4. To have the structure of history variables comparable to that of the standard model, we fitted the same history vector, that is, the baseline riskscore, the past exchange status, and the past distance from the NEP, $H_{i,t}=(B_i, E_{i,t-1}, D_{i,t-1})$. To have the models for outcome and censoring satisfy the compound exclusion restriction, we fitted

the link function introduced in Section 4. In particular, we parameterize links (4.1)–(4.3) as

$$\text{link}^{(S)}(H_{i,t}, t)\beta^{(S)}_{(h)} := \beta^{(S)}_{(b)}B_i + \beta^{(S)}_{(e-)}E_{i,t-1} + \beta^{(S)}_{(d-)}D_{i,t-1} + \beta^{(S)}_{(t)}t, \quad (5.1)$$

$$\text{link}^{(Y)}(D_{i,t}, S_{i,t}, H_{i,t}, t)\beta^{(Y)} := \beta^{(Y)}_{(0)} + \beta^{(Y)}_{(b)}B_i + \beta^{(Y)}_{(e-)}E_{i,t-1} + \beta^{(Y)}_{(d-)}D_{i,t-1} + \beta^{(Y)}_{(t)}t + \beta^{(Y)}_{(s)}S_{i,t} + \beta^{(Y)}_{(s,d)}E_{(S,D)}(S_{i,t}, D_{i,t}), \quad (5.2)$$

$$\text{link}^{(C)}(D_{i,t}, S_{i,t}, H_{i,t}, t)\beta^{(C)} := \beta^{(C)}_{(0)} + \beta^{(C)}_{(b)}B_i + \beta^{(C)}_{(e-)}\text{E}_{i,t-1} + \beta^{(C)}_{(d-)}D_{i,t-1} + \beta^{(C)}_{(t)}t + \beta^{(C)}_{(s)}S_{i,t} + \beta^{(C)}_{(s,d)}\text{E}_{(S,D)}(S_{i,t}, D_{i,t}). \quad (5.3)$$

With the parameterization (4.2) and (5.2), the causal effect that is the odds ratio between the top and bottom probabilities in (2.1) for principal strata $S_{i,t} = d$, $1 \le d \le d_{\max} - 1$, is equal to $\exp[\beta^{(Y)}_{(s,d)}\{\text{E}_{(S,D)}(d, 1) - \text{E}_{(S,D)}(d, d_{\max})\}]$, which, by multilevel monotonicity [Fig. 1(a)], equals $\exp\{\beta^{(Y)}_{(s,d)}(1 - 0)\} = \exp(\beta^{(Y)}_{(s,d)})$. Estimation, then, of this odds ratio causal effect can follow from joint estimation of the principal stratification model (5.1)–(5.3).

Fitting this model without further assumptions gives an estimated causal effect of 89% reduction in the odds of HIV attributable to exchange, or OR = .11, with a wide CI of (.002, 7.99). This means that, although, as shown in Section 3, the causal effect is theoretically estimable without further assumptions, and here is estimated substantially larger than in the standard method, some plausible additional structure is required to increase precision.

Because the principal stratification model accounts for any unmeasured risk for HIV between exchangers and nonexchangers, the 89% estimated reduction with this method compared to the 32% reduction with the standard method [Table 2(b)] suggests the hypothesis $\text{H}_{\text{mis|obs}}$ = "even after adjusting for the observed baseline risk factors, exchangers were at higher, thus unmeasured, risk for HIV than nonexchangers, independently of the act of exchange and of their distance from the NEP." Moreover, Table 2(a) [and comparison between Table 1, (a) and (b)] supports the hypothesis H_{obs} = "exchangers were at higher risk for HIV in the observed baseline factors compared to nonexchangers." By recalling that the principal stratum is the subject-specific unmeasured threshold distance for not exchanging, we can formalize the hypothesis that H_{obs} and $\text{H}_{\text{mis|obs}}$ are either both true or both false using our model, by

$$\beta^{(S)}_{(b)} \cdot \beta^{(Y)}_{(s)} \ge 0. \quad (5.4)$$

This condition still allows the standard assumption, that all confounders are measured, as a special case, but also allows for the possibility of some types of unmeasured confounding. Substantively in the NEP, (5.4) allows that in the variables $H_{i,t}$ we may not have measured the baseline sexual and drug risk behaviors in sufficient detail (e.g., due to underreporting), and, as a result, observed exchangers and nonexchangers may not be comparable in their risk for HIV seroconversion given $H_{i,t}$ and independently of exchange. Put another way, (5.4) means that we allow unmeasured confounding, but that it would be implausible if the unmeasured association in hypothesis $\text{H}_{\text{mis|obs}}$, after measuring 25 relevant baseline covariates, carries the opposite message from that of the measured association in H_{obs} about exchangers being at higher (or lower) risk than nonexchangers. Mathematically, (5.4) is not forced to be true or false in our general assumptions in Section 2 and so, by Section 3, (5.4) is testable in our framework. In the earlier fit of the model without constraint (5.4), we obtained estimates for $\beta^{(S)}_{(b)}$ and $\beta^{(Y)}_{(s)}$ equal to .14 and 1.07, respectively, and so the estimate for $\beta^{(S)}_{(b)} \cdot \beta^{(Y)}_{(s)}$ equals $.15 > 0$. Therefore, our data can provide no evidence against (5.4) and, so we take our model to be (5.1)–(5.4).

Table 3. Results From the Principal Stratification Model of Section 5.3 on Exchanging Needles at the NEP and on HIV Seroconversion

	Estimand	Estimate	95% CR[b]
(a)	Odds ratio of higher versus lower principal stratum[a] S for 1 standard deviation increase in baseline risk score B: $\exp(\beta^{(S)}_{(b)})$	1.15	(1.11, 1.20)
(b)	Odds of HIV seroconversion under fixed exchange for higher versus lower principal stratum[a] S, given fixed baseline risk score B: $\exp(\beta^{(Y)}_{(s)})$	2.92	(.99, 96.36)
(c)	Odds ratio of HIV seroconversion for close versus far distance from the NEP and attributable to needle exchange[c]: $\exp(\beta^{(Y)}_{(s,d)})$	.11	(.0003, 2.23)

[a] A subject's principal stratum is the closest distance to place the NEP beyond which that subject would not exchange at it.

[b] Joint confidence region.

[c] Conditionally on baseline risk score B and principal stratum S.

To obtain inference for this model, the MLE's of $L(\beta)$ remain unchanged because they satisfy (5.4). To incorporate the implications of (5.4) on the uncertainty about the causal effect $\exp(\beta^{(Y)}_{(s,d)})$, we take the ellipsoidal three-dimensional joint 95% confidence region R' for $\beta^{(Y)}_{(s,d)}$ and the two estimands, $\beta^{(S)}_{(b)}$ and $\beta^{(Y)}_{(s)}$, involved in (5.4), and find the subset R of R' that satisfies (5.4). The resulting region R, then, is a 95% joint confidence region for the three estimands.

These results are given in Table 3. The estimated effect of the NEP that is attributable to exchanging versus not exchanging needles is a reduction by 89% in HIV seroconversion, as in the fit of this model without (5.4). The point estimate, thus, means that the data point to a larger benefit of the NEP, compared to the standard method. The uncertainty about the effect (OR = .11, 95% joint CR: .00, 2.23) is larger than that of the standard approach [Table 2(b)]. This wider uncertainty fairly reflects that the new method accounts for possible confounding remaining in the standard approach because exchangers and nonexchangers are different mixtures of the partially unmeasured principal strata.

6. DISCUSSION

We proposed a method based on principal stratification for evaluating studies where (1) subjects take a longitudinal treatment, whose transient effect on a time-to-event outcome is of interest, but where this treatment is not directly controlled; (2) the subject's follow-up time is not directly controlled; and (3) the study directly controls another longitudinal factor that

can affect both exposure to the treatment of interest as well as follow-up time. The new method based on principal strata has three advantages over existing methods. First, the method based on principal strata can better address studies with coexistence of partially controlled factors such as exposure and censoring of outcomes. Second, the method of instrumental variables is, in principle, based on estimation of coefficients in additive equations with error terms, and so can only estimate causal effects that happen to be expressible in terms of such equations. In contrast, the method based on principal strata estimates causal effects whose definition and interpretation does not require parametric models. Rather, the role of parametric models is in assisting estimation. Finally, the method based on principal strata can make the assumptions explicit. Making assumptions such as those of Section 2.2 explicit can allow researchers to evaluate their plausibility.

The notation we used for a subject's potential outcomes, exposure, and censoring behaviors as functions of that subject's level of the controlled factor D also embodies the implicit assumption that those values are not a function of other subjects' levels of the controlled factor, namely, the no-interference assumption (Cox 1958, p. 19; Rubin 1978). Such an assumption is typically made, and relaxing it leads to many more potential outcomes and principal strata, so that inferences without it would require additional assumptions. For example, as pointed out by a reviewer, interference in the NEP between a couple of cohabiting drug users can exist if they are both recruited in the study and their network for sharing needles is limited to only the two of them. Then the exchanging behavior at the NEP for one member of the couple is expected to be very influential on the potential outcome of the other member, even when the latter member does not exchange at the NEP. On the other hand, interference is not expected to be a serious concern if any impact of the study on its subjects does not result in a big change on the larger network of drug users with which the study drug users interact. Inference using this assumption in this study is relevant because (a) the study subjects are only a subset of the larger network of drug users with whom they interact; (b) within the study, and at any time, only a small fraction of subjects exchange at the NEP; and (c) small deviations from the assumption would mean that the effect of distance attributable to exchange as estimated in Section 5 is a conservative estimate of the target impact that can be realized if the larger part of the community of drug users is to visit and exchange at the NEP. Nevertheless, our framework can, in principle, allow interference by allowing the subject-specific potential values of outcomes, exchange, and censoring at a given time to be also a function of the subjects' collective behaviors at previous times. Such models require additional assumptions to identify causal effects and are of interest in further work.

The described method also assumes observations across subjects are independent, which, if not true unconditionally, becomes more realistic when conditioning on important covariates. However, direct fitting of a large number of covariates to address possible spatial clustering can be avoided by modeling instead a correlation structure unconditionally on the covariates. In data of simpler structure than that described here, work related to this issue is discussed in Frangakis, Rubin, and Zhou (1998, 2002) and Korhonen, Loeys, Goetghebeur, and Palmgren (2000), and generalizations to the longitudinal data structure of Section 2 can be useful.

Principal stratification can also be used in conjunction with other methods, for example, propensity scores (Rosenbaum and Rubin 1983). With settings such as those considered here, for example, for the NEP, estimating propensity scores is conceivable for the distribution of the controlled factor, or of the partially controlled observed exposured or of censoring. It is important to note, however, that the full advantage of using propensity scores comes mostly in simpler settings where assignment of a binary factor is assumed ignorable, given observed variables. Therefore, such full advantage of using propensity scores is not likely achievable in our setting where the controlled factor is generally multilevel and where exposure and censoring are allowed to be nonignorable. Nevertheless, further study of the role of propensity scores, for example, for dimension reduction without introducing bias, in settings such as ours is also of interest.

With better methods of analysis, there is also a need for better retrospective and prospective designs. An example of the former in the NEP could be: Choose the known cases of HIV and, from the controls of the study (at each time point), select only the most informative ones, by means of appropriate measures of covariates, of controlled distance, and of partially controlled exposure. The motivation for such approaches is that, although we would be using reduced data, we would have more accurate estimates than if we were using the full data if model extrapolation in the latter case can introduce large bias relative to gains in efficiency. In this case one can use the conditional likelihood that is induced by the matching scheme on the unconditional likelihood of models such as those of Section 4. An important issue, however, in such a problem is to select informative matching schemes, given the different role of the controlled and partially controlled factors in the study. For prospective designs, methods are needed to develop guidelines for the controlled factor, for example, in the NEP, about how many sites, and at what locations, should be used to achieve balance between, on the one hand, current practical and ethical concerns and, on the other hand, precision of estimation of effects of a program that can be beneficial in the future for the larger community. Work on prospective designs that anticipate noncompliance discussed by Jo (1999) and Frangakis and Baker (2001) provides some preliminary directions that are relevant to the role of prospective designs in such more general settings.

Regarding evaluation of the Baltimore NEP, the method that uses principal stratification points to a substantially larger benefit of the NEP in reducing HIV transmission than the standard method, although the results for both methods are associated with considerable uncertainty. More important than the results of each method alone is that the different methods give consistent results and in relative magnitudes that are explainable if the NEP does lower HIV transmission and also attracts subjects who are at higher risk. Such consistency between the two different perspectives provides more confidence on the effectiveness of the NEP, although continued follow-up can provide more conclusive information.

Documentation and software for the methods described here are available online at *http://biosun01.biostat.jhsph.edu/~cfrangak/papers/ps.html*.

[Received XXXX. Revised XXXX.]

REFERENCES

Angrist, J. D., and Imbens, G. W. (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables" (with discussion), *Journal of the American Statistical Association*, 91, 444–472.

Baker, S. G. (1998), "Analysis of Survival Data From a Randomized Trial With All-or-None Compliance: Estimating the Cost-Effectiveness of a Cancer Screening Program," *Journal of the American Statistical Association*, 93, 929–934.

——— (2000), "Analyzing a Randomized Cancer Prevention Trial With a Missing Binary Outcome, an Auxiliary Variable, and All-or-None Compliance," *Journal of the American Statistical Association*, 95, 43–50.

Baker, S. G., and Lindeman, K. S. (1994), "The Paired Availability Design: A Proposal for Evaluating Epidural Analgesia During Labor," *Statistics in Medicine*, 13, 2269–2278.

——— (2001), "Rethinking Historical Controls," *Biostatistics*, 2, 383–396.

Baker, S. G., Lindeman, K. S., and Kramer, B. S. (2001), "The Paired Availability Design for Historical Controls," *BMC Medical Research Methodology*, 1–9.

Barnard, J., Frangakis, C. E., Hill, L., and Rubin, D. B. (2002), "School Choice in NY City: A Bayesian Analysis of an Imperfect Randomized Experiment," in *Case Studies in Bayesian Statistics*, New York: Springer-Verlag, pp. 3–97.

Bowden, R. J., and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge, U.K.: Cambridge University Press.

Bruneau, J., Lamothe, F., and Franco, E. (1997), "High Rates of HIV Infection Among Injection Drug Users Participating in Needle Exchange Programs in Montreal: Results of a Cohort Study," *American Journal of Epidemiology*, 146, 994–1002.

Card, D. (1993), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," Paper 4483, National Bureau of Economic Research.

Cochran, W. G. (1957), "Analysis of Covariance: Its Nature and Uses," *Biometrics*, 13, 261–281.

Cox, D. R. (1958), *Planning of Experiments*, New York: Wiley.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman & Hall.

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society*, Ser. B, 41, 1–31.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Frangakis, C.E., and Baker, S.G. (2001), "Compliance Sub-sampling Designs for Comparative Research: Estimation and Optimal Planning," *Biometrics*, 57, 899–908.

Frangakis, C. E., and Rubin, D. B. (1997), "A New Approach to the Idiosyncratic Problem of Drug-Noncompliance With Subsequent Loss to Follow-Up," in *Proceedings of the Biometrics Section, American Statistical Association*, pp. 206–211.

——— (1999), "Addressing Complications of Intention-to-Treat Analysis in the Combined Presence of All-or-None Treatment-Noncompliance and Subsequent Missing Outcomes," *Biometrika*, 86, 365–379.

——— (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 20–29.

Frangakis, C. E., Rubin, D. B., and Zhou, X. H. (1998), "The Clustered Encouragement Design," in *Proceedings of the Biometrics Section, American Statistical Association*, pp. 71–79.

——— (2002), "Clustered Encouragement Design With Individual Noncompliance: Bayesian Inference and Application to Advance Directive Forms" (with discussion), *Biostatistics*, 3, 147–164.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–411.

Hagan, H., McGough, J. P., Thiede, H., Weiss, N. S., Hopkins, S., and Alexander E. R. (1999), "Syringe Exchange and Risk of Infection With Hepatitis B and C Viruses," *American Journal of Epidemiology*, 149, 203–213.

Hernan, M. A., Brumback, B., and Robins, J. M. (2000), "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men," *Epidemiology*, 11, 561–570.

Imbens, G. W., and Rubin, D. B. (1994), "Causal Inference With Instrumental Variables," Discussion Paper 1676, Cambridge, MA: Harvard Institute of Economic Research.

——— (1997), "Bayesian Inference for Causal Effects in Randomized Experiments With Noncompliance," *The Annals of Statistics*, 25, 305–327.

Jo, B. (1999), "Power to Detect Intervention Effects in Randomized Trials With Noncompliance," technical report, UCLA, Graduate School of Education and Information Studies.

Korhonen, P., Loeys, T., Goetghebeur, E., and Palmgren, J. (2000), "Vitamin A and Infant Mortality: Beyond Intention-to-Treat in a Randomized Trial," *Lifetime Data Analysis*, 6, 107–121.

McClellan, M., McNeil, B. J., and Newhouse, J. P. (1994), "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables," *Journal of the American Medical Association*, 272, 859–866.

Murphy, S. A., van der Laan, M., Robins, J., and CPPRG (2001), "Marginal Mean Models for Dynamic Regimes," *Journal of the American Statistical Association*, 96, 1410–1423.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9," translated in *Statistical Science* (1990), 5, 465–480.

Robins, J. M. (1986), "A New Approach to Causal Inference in Mortality Studies With Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393–1512.

Robins, J. M., and Greenland, S. (1994), "Adjusting for Differential Rates of Prophylaxis Therapy for PCP in High-Versus Low-Dose AZT Treatment Arms in an AIDS Randomized Trial," *Journal of the American Statistical Association*, 89, 737–749.

Robins, J. M., Greenland, S., and Hu, F.-C. (1999), "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome" (with discussion), *Journal of the American Statistical Association*, 94, 687–712.

Rockwell, R., Des Jarlais, D. C., Friedman, S. R., Rerlis, T. E., and Paone, D. (1999), *AIDS Care*, 4, 437–442.

Rosenbaum, P. R. (1984), "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society*, Ser. A, 147, 656–666.

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

——— (1977), "Assignment to a Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1–26.

——— (1978), "Bayesian Inference for Causal Effects," *The Annals of Statistics*, 6, 34–58.

——— (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.

——— (1996), "Multiple Imputation After 18+ Years" (with discussion), *Journal of the American Statistical Association*, 91, 473–489.

——— (1998), "More Powerful Randomization-Based p-Values in Double-Blind Trials With Noncompliance," *Statistics in Medicine*, 17, 371–385.

Rubin, D. B., and Frangakis, C. E. (1999), Comment on "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome," by J. M. Robins, S. Greenland, and F.-C. Hu, *Journal of the American Statistical Association*, 94, 702–704.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman & Hall.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999), "Adjusting for Nonignorable Drop-out Using Semiparametric Nonresponse Models" (with discussion), *Journal of the American Statistical Association*, 94, 1096–1146.

Sommer and Zeger (1991).

Strathdee, S. A., Celentano, D. D., Shah, N., Lyles, C., Stambolis, V. A., Macalino, G., Nelson, K., and Vlahov, D. (1999), "Needle-Exchange Attendance and Health Care Utilization Promote Entry Into Detoxification," *Journal of Urban Health*, 76, 448–460.

Vlahov, D., Anthony, J. C., Munoz, A., et al. (1991), "The ALIVE Study. A Longitudinal Study of HIV Infection Among Injection Drug Users: Description of Methods," *NIDA Research Monograph*, 107, 75–100.

Vlahov, D., Junge, B., Brookmeyer, R., et al. (1997), "Reduction in High-Risk Drug Use Behaviors Among Participants in the Baltimore Needle Exchange Program," *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 16, 400–406.

Wald, A. (1940), "The Fitting of Straight Lines When Both Variables Are Measured With Error," *Annals of Mathematical Statistics*, 49, 346–350.